DOCUMENT RESUME

ED 201 679 TM 810 278

AUTHOR Frechtling, Joy A.
TITLE The Use and Misuse of Norm Referenced Test Scores for Decision Making at the Local School Level: A School District's Perspective.
PUB DATE Apr 81
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Aptitude Tests; *Decision Making; Elementary Secondary Education; *Norm Referenced Tests; *School Districts; *Testing Problems; *Test Interpretation
IDENTIFIERS *Test Use

ABSTRACT

The traditional use of scores derived from standardized, norm-referenced achievement tests is examined as to its role in a local school system's efforts to monitor itself and to examine how well individual students, schools and the overall system are functioning. Several questions regarding the legitimacy of well-established practices involving test use are raised. These questions include: (1) how much and what kind of information do norm-referenced standardized tests really provide for program managers and school administrators, and (2) what are the real differences between an achievement test and an ability test in what they measure and how they should be used? It is concluded that the use of standardized tests should be closely examined, and that alternative methods are needed for measuring achievement expectancy of students as well as schools. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED201679

The Use and Misuse of Norm Referenced Test Scores for Decision Making at the Local School Level: A School District's Perspective

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Joy A. Frechtling
Montgomery County Public Schools

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

J. Frechtling

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

April, 1981

Paper presented at the annual meeting of the American Educational Research Association
Los Angeles, California

TM 810278

Background

Testing plays a major role in a local school system's efforts to monitor itself and to examine how well individual students, schools, and the overall system is functioning. In fact, many assessors of quality education use test results as the major barometer of success or failure of public education and place the number derived from a series of well filled bubbles far above any other index of how well children are being taught. In the school system in which I work, Montgomery County Maryland, we (those people in charge of administering and interpreting test data) have for the past two years been doing some serious soul searching regarding standardized test scores and the information they provide. We have looked at the use to which the tests have been put and come to pose some very serious questions regarding the legitimacy of well established practices.

The questions we have been asking fall into two general categories. First, how much and what kind of information do norm referenced standardized tests really provide for program managers and school administrators? And, second, what are the real differences, if any, between an achievement test and a so-called "abilities" test in what they measure and how they should be used?

Uses of Tests

In order to understand more fully our concerns and why these questions have arisen, it may be helpful to look at the ways in which test data have been used in Montgomery County. In all likelihood, our practices can be considered

reasonably typical of practices in major school systems nationwide. Standardized norm referenced achievement tests are given as part of a statewide testing program. Achievement tests are given to all students in Grades 3, 5, and 8. Until this year, administration of an abilities test was also required at the same grade levels. Now, in the eyes of the state, abilities testing is considered optional.

The traditional uses of scores derived from these tests fall into four general categories -- evaluating the skills and needs of individual students, evaluating schools, evaluating instructional programs, and evaluating the school system. These are discussed briefly below:

1. Evaluating the skills and needs of individual students. On the individual level, norm referenced test data have had several uses.

   - Individual diagnosis and programming - to provide staff with information which supplements grades and professional judgement, to be used in determining individual needs and suggesting activities from which a student might benefit. Comparisons between achievement and abilities test performance have been considered by staff to be an important indicator of satisfactory progress, "over" or "under" achievement.

   - Communication with parents - to inform parents about the educational attainments of their children. Comparisons between achievement and abilities test scores have provided one indicator of student motivation, as well as the degree to which a school is meeting a student's needs.

- o Screening for special programs - to help staff in selecting students for special programs, such as gifted and talented and advanced placement courses. Abilities test performance, in particular has until very recently, played a major role in deciding among students who otherwise appear to have equivalent qualifications.

- o Grouping - to place students in special classes or groups within classes which differ in content, level, or pace of instruction.

2. Evaluating schools. There are two general ways in which standardized test data have been used to evaluate schools.

   - o Assessing individual school performance - to determine whether or not individual schools are providing quality instruction which meets the needs of its student population. As in the case of individual diagnosis, comparisons have frequently been made between school ability and achievement test scores to determine whether or not the school, and by inference the principal, was functioning at, above, or below what might be expected given the ability level of the enrollees.

   - o Comparing performance among schools - to determine which schools are the best and to rank schools vis a vis each other in terms of their academic accomplishments. (The real estate agent's dream). Sometimes abilities test scores have been used as a kind of "control variable" in making such comparisons.

3. Evaluating the effectiveness of instructional programs - to determine whether different approaches to teaching a subject, such as reading, are differentially effective. As described above in evaluation designs abilities test scores have sometimes been used as a covariate.

4. Evaluating school system quality - to determine whether or not the system as a whole is providing a quality educational program for its students. Here, differences between systemwide abilities and achievement test scores (or lack there of) have again been used as a standard against which to measure performance. In many cases, in making such comparisons the district superintendent is the benefactor of either the praise or the blame which accrues from this activity.

While this list is lengthy, it is likely that it is under rather than over inclusive. I think its fair to say that information from standardized norm referenced tests is used extremely widely and effects decision making at all levels of local educational systems.

The Problem

Why do we question these uses? Three major concerns will be discussed here. And, up front I want to point out that, to some extent, these concerns stem from misinterpretation or misuse of test data by well meaning believers in the power of the "objective" quantitative approach and are not fault of the tests themselves or the test developers/publishers. To differing degrees, however, I feel it is fair to say that the test marketers and their written materials encourage such uses.

The first concern is that many of these practices involve using data from tests primarily intended to provide data on groups to make decisions about individual students. These tests are not only group administered but designed to be most reliable when used to measure group performance. We question whether or not the scores for individuals are sufficiently accurate for the uses to which they are being put. For example, given the standard error of the test, can one really say that a student with a score at the 95th percentile is superior to one with a score at the 93rd or even 90th percentile? Can a cut score be justified?

The second concern is whether there is in fact any useful difference between what is measured by ability and achievement tests and whether the practice of comparing scores on the two measures, as was mentioned in several of the examples above, makes sense. This is clearly not a new issue in the area of test usage. It is one that remains hotly debated, however, as the recommendation that abilities tests be used as a standard against which to measure achievement test performance continues to be made. This has become a very emotionally charged area. Further, there is so much disagreement among experts that the potential for debate seems almost endless. For example, at a work session before our Board of Education last spring out of ten experts our department assembled we managed to have five who endorsed the practice and five who did not. This hardly provided us with convincing support for rejecting a practice with a long history in our county. Nonetheless, we question whether or not the tests differ more than two so-called achievement tests and are very uncomfortable with setting one as the basic standard against which to examine performance on the other.

The third concern is that in general we feel that the public and educators tend to overrate the information provided by standardized test scores. There is something seductive about the apparent simplicity and objectivity of a number derived from a paper and pencil test. However, the kinds of decisions for which they are used are in fact very complicated. Standardized test scores are only one of many factors which should be taken into account in drawing conclusions. Unfortunately, this is all too infrequently the case. The publication of school by school test scores always makes headlines in the local papers and leads parents and the public to draw rapid and sometimes quite inaccurate inferences about school/principal performance.

Solutions

We have found that pointing out these problems and raising these questions does not readily lead to consensus or modification of practice. For policy makers (Board of Education members, administrators, etc) it is not sufficient to say that a practice is invalid, especially a practice that has proven useful. An alternative solution must be offered.

We have spent a good deal of time over the last year trying to find some alternative solutions. If I had to give our efforts a grade, I think a "B" would be considered fair. We've done pretty well, for example, in the area of student selection for special programs. Strict ranking by test scores is now

officially discouraged (although not eliminated in practice) and the importance placed on scores on so-called abilities tests over other measures is decreased. In time we may even see more emphasis placed on work samples and other more direct measures of skill level in a specific area.

We have introduced a different way to look at whether or not schools are effective. Specifically, we now do comparisons of longitudinal data for students tested in the same school at more than one grade level (e.g. in the third and fifth grade for elementary schools). We look to see whether there is any "substantial" difference in test performance (composite or subtest) for the group between the two test periods. To reach this judgment we assume that all other things being equal students would be expected to rank similarly at the two test points. In other words, the best predictor of future performance is current performance. If performance across the grades differs by a specified amount ($\pm$7 NCE points or a 1/3 of a standard deviation from the county trend), it is considered an indicator that the school may be especially effective (or ineffective) in the area the test is measuring. It should be noted that in making judgements about the performance of individual schools the countywide trend is considered because it is important to guard against attributing to a school strengths or weaknesses that in reality relate to the countywide curriculum. Thus, if the county trend were to increase from 80 NCEs to 82 NCEs a gain of a 2 NCEs, a school would have to show an increase of 10 NCEs to be considered effective.

This way of assessing school effectiveness seems to us to be far better than comparing performance on achievement and so-called abilities tests. However,

we are aware that it too has shortcomings. One could argue that ± 7 NCEs is a rather arbitrary figure and that there is no convincing basis for selecting that criteria over others. Regression to the mean may be occurring for our extreme scoring schools, and we are not quite sure as yet how to take this into account. We are not convinced that regression analyses totally solve our problem. In addition under these criteria if a principal wanted to manipulate the system, she could place her weakest teachers in the third grade and her strongest teachers in the fourth and fifth grades.

For these reasons we stress that this analysis provides an indicator of whether a school may be more or less effective and trends across multiple years should be examined. We try to emphasize that it is a way of flagging schools for further study by professionals more familiar with curricula and instruction. We are encouraged, however, by the potential of this approach especially since some of the schools which are flagged include ones not typically cited as being outstanding where traditional school ranking methods (ranking according to a single year's performance) are used. Specifically, for the first time, schools in the highest SES areas are showing up as having academic problems and schools in relatively lower SES areas are identified as having noteworthy programs.

We have not totally succeeded in convincing people that this approach is better than one which compares achievement to ability and it clearly is not a direct substitute. It does not answer the question of whether a school or a child is doing as well as he/she should. However, we are unconvinced that we have or ever have had a valid measure of this expectation. Unfortunately, this opinion is not shared by some very important policy makers.

Conclusions

School districts continue to struggle with the problem of interpreting and using standardized test data. While scores on standardized norm referenced tests can be very helpful when used appropriately as a decision-making tool, the same data, when misused, can do extreme damage to individuals and institutions. While we want to continue to examine the use of standardized tests for all the purposes described earlier we are especially concerned about the use of abilities tests, and answering questions regarding what they measure, how, and, if, they should be used in ways which differ from uses of achievement tests. We are also looking for suggestions for alternative methods for measuring expectancy or how well a school or child "should" be doing if such exist. Any light that this panel sheds on these issues will be greatly appreciated.

4079A/jal